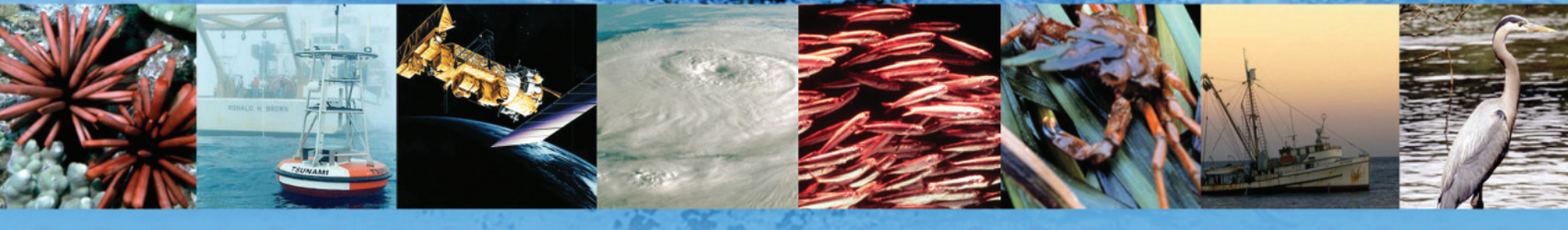
National Oceanic and Atmospheric Administration

Report to Congress on Data and Information Management

2007



THE SECRETARY OF COMMERCE
Washington, D.C. 20230

July 28, 2008

The Honorable Daniel K. Inouye
Chairman, Committee on Commerce,
Science and Transportation
United States Senate
Washington, DC 20510

Dear Mr. Chairman:

I am pleased to transmit the National Oceanic and Atmospheric Administration's (NOAA) *Report to Congress on Data and Information Management 2007*. This report was produced in compliance with Public Law 102-567, Section 106.

Within the Department of Commerce, NOAA's mission is to understand and predict changes in the Earth's environment and to conserve and manage coastal and marine resources to meet our Nation's economic, social, and environmental needs. NOAA met this mission in 2007 through international cooperation, data product creation and improvement, and environmental stewardship. NOAA continues to be an international leader in optimizing the way integrated environmental observations and information are stored, shared, and used to benefit the world. As a result, our environmental information products are valuable national and international resources. We are now focused on increasing access to environmental data and information for users around the globe.

Throughout the 1990s and into the new century, the NOAA archives have experienced significant growth. The Internet has facilitated an unprecedented increase in the number of users who request archive access to NOAA environmental data and information. These archive and usage trends are expected to continue for the foreseeable future. Over the past decade, NOAA has made great strides with respect to its data management mission. This report updates our progress. It also defines the course of action that will enable us to continue to preserve and provide access to important environmental data for the Nation.

If you have any questions, please contact me or Nathaniel F. Wienecke, Assistant Secretary for Legislative and Intergovernmental Affairs, at (202) 482-3663.

Sincerely,

Carlos M. Gutierrez

Enclosure

THE SECRETARY OF COMMERCE
Washington, D.C. 20230

July 28, 2008

The Honorable Bart Gordon
Chairman, Committee on Science and Technology
U.S. House of Representatives
Washington, DC 20510

Dear Mr. Chairman:

I am pleased to transmit the National Oceanic and Atmospheric Administration's (NOAA) *Report to Congress on Data and Information Management 2007*. This report was produced in compliance with Public Law 102-567, Section 106.

Within the Department of Commerce, NOAA's mission is to understand and predict changes in the Earth's environment and to conserve and manage coastal and marine resources to meet our Nation's economic, social, and environmental needs. NOAA met this mission in 2007 through international cooperation, data product creation and improvement, and environmental stewardship. NOAA continues to be an international leader in optimizing the way integrated environmental observations and information are stored, shared, and used to benefit the world. As a result, our environmental information products are valuable national and international resources. We are now focused on increasing access to environmental data and information for users around the globe.

Throughout the 1990s and into the new century, the NOAA archives have experienced significant growth. The Internet has facilitated an unprecedented increase in the number of users who request archive access to NOAA environmental data and information. These archive and usage trends are expected to continue for the foreseeable future. Over the past decade, NOAA has made great strides with respect to its data management mission. This report updates our progress. It also defines the course of action that will enable us to continue to preserve and provide access to important environmental data for the Nation.

If you have any questions, please contact me or Nathaniel F. Wienecke, Assistant Secretary for Legislative and Intergovernmental Affairs, at (202) 482-3663.

Sincerely,

Carlos M. Gutierrez

Enclosure

THE SECRETARY OF COMMERCE
Washington, D.C. 20230

July 28, 2008

The Honorable Ralph M. Hall
Ranking Minority Member
Committee on Science and Technology
U.S. House of Representatives
Washington, DC 20510

Dear Representative Hall:

I am pleased to transmit the National Oceanic and Atmospheric Administration's (NOAA) *Report to Congress on Data and Information Management 2007*. This report was produced in compliance with Public Law 102-567, Section 106.

Within the Department of Commerce, NOAA's mission is to understand and predict changes in the Earth's environment and to conserve and manage coastal and marine resources to meet our Nation's economic, social, and environmental needs. NOAA met this mission in 2007 through international cooperation, data product creation and improvement, and environmental stewardship. NOAA continues to be an international leader in optimizing the way integrated environmental observations and information are stored, shared, and used to benefit the world. As a result, our environmental information products are valuable national and international resources. We are now focused on increasing access to environmental data and information for users around the globe.

Throughout the 1990s and into the new century, the NOAA archives have experienced significant growth. The Internet has facilitated an unprecedented increase in the number of users who request archive access to NOAA environmental data and information. These archive and usage trends are expected to continue for the foreseeable future. Over the past decade, NOAA has made great strides with respect to its data management mission. This report updates our progress. It also defines the course of action that will enable us to continue to preserve and provide access to important environmental data for the Nation.

If you have any questions, please contact me or Nathaniel F. Wienecke, Assistant Secretary for Legislative and Intergovernmental Affairs, at (202) 482-3663.

Sincerely,

Carlos M. Gutiérrez

Enclosure

NOAA's Vision and Mission Goals

NOAA'S VISION

An informed society that uses a comprehensive understanding of the role of the oceans, coasts, and atmosphere in the global ecosystem to make the best social and economic decisions

NOAA'S MISSION

To understand and predict changes in the Earth's environment, and conserve and manage coastal and marine resources to meet our Nation's economic, social, and environmental needs

NOAA'S MISSION GOALS

- Protect, restore, and manage the use of coastal and ocean resources through an ecosystem approach to management
- Understand climate variability and change to enhance society's ability to plan and respond
- Serve society's needs for weather and water information
- Support the Nation's commerce with information for safe, efficient, and environmentally sound transportation
- Provide critical support for NOAA's mission

Foreword

It is my pleasure to present the National Oceanic and Atmospheric Administration's (NOAA's) *Report to Congress on Data and Information Management 2007*. Within the Department of Commerce, NOAA's mission is to understand and predict changes in the Earth's environment and conserve and manage coastal and marine resources to meet our Nation's economic, social, and environmental needs. NOAA advanced this mission in 2007 through international and interagency cooperation, data product creation and improvement, and environmental stewardship. NOAA develops and applies environmental science and technology to the problems our Nation faces in a changing world.

NOAA has increased international engagements to encourage integrated environmental observations and promote accessible information that is stored, shared, and used for the benefit of the world. Our environmental information products are valuable national and international resources. The "One NOAA" Initiative has focused attention on increasing access to environmental data and information for users both at home and abroad.

In addition, NOAA has improved products that provide environmental information and resources to scientists, researchers, and the public. This information also provides policy makers with the tools to implement informed decisions regarding our environment. Indeed, NOAA touches the lives of Americans and foreign citizens every day.

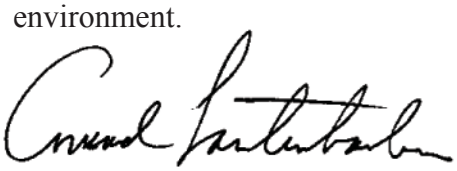
Carlos M. Gutierrez
Secretary of Commerce

Preface

Data and information management are vital components of the National Oceanic and Atmospheric Administration's (NOAA's) continuing efforts to maintain world-class environmental science. By advancing data and information management, we can be more effective global environmental stewards and excel at providing the scientific community with advanced tools, products, and opportunities to help benefit society and the world around us. NOAA's *Report to Congress on Data and Information Management 2007* lays out our plans to address the challenges we face. The report also demonstrates that NOAA is making progress toward a comprehensive data management architecture, safeguarding NOAA's priceless data assets.

Our national infrastructure capabilities and expertise contribute to our leadership role in the intergovernmental Group on Earth Observations (GEO). Understanding our global environment requires a global effort, and we are continuing to build a global network to share data and information. Management of data and information on a global scale is being addressed as part of the Global Earth Observation System of Systems (GEOSS), which is being developed by GEO.

Our users are more sophisticated and more dependent on us than in the past, and thus time critical information delivery is essential. Together we can employ NOAA's environmental data and information resources to enable timely responses to threats against life, property, and the environment.



Conrad C. Lautenbacher, Jr.
Vice Admiral, U.S. Navy (Ret.)
Under Secretary of Commerce for Oceans and Atmosphere and NOAA Administrator

Table of Contents

Executive Summary

This report describes the current state of data management within the National Oceanic and Atmospheric Administration (NOAA), as well as NOAA's plans for continued improvement in meeting its mandate to collect, preserve, and distribute data of the highest quality. This is the latest in a series of biennial reports to Congress on NOAA data and information management, prepared in response to congressional direction contained in Public Law 102-567, Section 106: Data and Information Systems. The two main parts of this report reflect the two principal reporting requirements contained in the Public Law: (1) "an assessment of the adequacy of the environmental data and information systems of the National Oceanic and Atmospheric Administration" and (2) a description of "a comprehensive plan . . . to modernize and improve" those systems.

Following the approach taken in 2005, each of NOAA's crosscutting Goal Teams—Ecosystems, Climate, Weather and Water, and Commerce and Transportation—was asked to assess program status with respect to the full range of data management activities, from observation acquisition and transmission to long-term archival access. The results show that in most cases, NOAA's Goal Teams are making incremental progress toward collecting high-quality data, describing them carefully, maintaining them securely over the long term, and effectively and efficiently integrating the many existing data systems. NOAA reports successes in addressing the five data management challenges in the 2005 assessment.

NOAA has made substantial progress in planning for an integrated "One NOAA" data management architecture. The cornerstone of this planning effort is the Global Earth Observation Integrated Data Environment (GEO-IDE), which lays out a framework enabling the effective and efficient integration of NOAA's many existing systems and serves as a guideline for future data system development efforts. The objectives and approach of GEO-IDE are consistent with the Group on Earth Observations (GEO) international partnership and support the Strategic Plan for the U.S. Integrated Earth Observation System (IEOS), the U.S. contribution to the Global Earth Observation System of Systems (GEOSS). In the FY 2008 budget request, NOAA proposes building an Initial Operating Capability (IOC) for its portion of the ocean component of IEOS, the Integrated Ocean Observing System (IOOS), which includes development of a data integration framework for a limited subset of ocean variables.

Report to Congress on Data and Information Management 2007

I. Introduction

NOAA's environmental services are a vital national resource, all enabled by effective data and information management.

This report is the eighth in a series of Data and Information Management Reports to Congress since the public law establishing the reporting requirement was passed in 1992. The current report is similar in format to the 2005 report, allowing for straightforward analysis of successes and challenges. Overall, the picture is one of continued—though measured—progress in achieving NOAA's vision of a fully integrated enterprise data system.

The NOAA National Data Centers (NNDCs) have the unique responsibility for the long-term management and stewardship of the bulk of NOAA's data—in addition to environmental data collected by other Federal agencies, countries, and research programs—for use in addressing today's and tomorrow's environmental issues. Also, there are numerous distributed NOAA Centers of Data responsible for the management of data sets developed in the process of fulfilling their particular environmental missions and operational responsibilities. Together with the infrastructure required to transport data from place to place, these National Data Centers and Centers of Data comprise the core of the NOAA data management enterprise, which is administered under the same reporting structure as in 2005. The NOAA Observing Systems Council (NOSC) oversees observing systems and data and information management and planning in consultation with the Chief Information Officer (CIO) (see Figure 1). The Data Management Committee (DMC) reports to the NOSC and coordinates the development and implementation of NOAA's data management policy. The Data Management Integration Team (DMIT) is composed of NOAA data management specialists and includes representatives from all NOAA line offices and goal

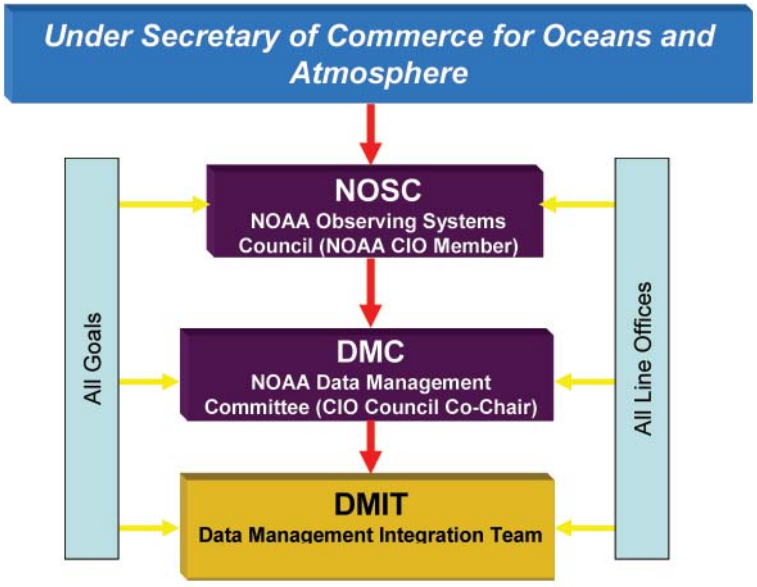


Figure 1. Reporting Structure for NOAA Data and Information Management.

areas and the office of the Chief Information Officer. The DMIT is charged with providing advice and expertise and with contributing to the integration of data management across NOAA and IOOS. In 2007, NOAA filled the newly created position of Data Management Integration Architect and DMIT chair.

NOAA continues to participate actively in USGEO, a subcommittee of the Committee on Environment and Natural Resources under the President's National Science and Technology Council. NOAA is a leader and major contributor to the work of the international GEO. NOAA is restructuring to better coordinate its role in developing IOOS, including oversight of the IOOS Data Management and Communications activities.

Since their establishment in 2004, the NOAA Mission Goal Teams have continued to evolve and are achieving success in driving the integration of NOAA data products and services across the traditional Line Office boundaries. Beginning in late 2006, NOAA instituted additional measures to facilitate integrated product and service delivery at the regional level. NOAA Regional Teams will seek to identify and apply the full range of NOAA capabilities, both within and across regions, to design the best geographically specific solutions for customers.

II. Assessment of Current Capabilities

Our assessment is governed by the provisions of Public Law 102-567, Section 106 (c) (1) which states:

"In conducting such an assessment, the Secretary shall take into consideration the need to -
(A) provide adequate capacity to manage, archive, and disseminate environmental data and information collected and processed, or expected to be collected and processed, by the National Oceanic and Atmospheric Administration and other appropriate departments and agencies;
(B) establish, develop, and maintain information bases, including necessary management systems, which will promote consistent, efficient, and compatible transfer and use of data;
(C) develop effective interfaces among the environmental data and information systems of the National Oceanic and Atmospheric Administration and other appropriate departments and agencies;
(D) develop and use nationally accepted formats and standards for data collected by various national and international sources; and
(E) integrate and interpret data from different sources to produce information that can be used by decision makers in developing policies that effectively respond to national and global environmental concerns."

This 2007 assessment of the adequacy of NOAA's environmental data and information systems is an update to the previous assessment, completed in 2005, and follows that report's structure and content. Information was acquired on data and information management across each of the NOAA Goals as described in the NOAA Strategic Plan. In 2005, the primary vehicle to gather information was a questionnaire. This year the assessment relies upon information that was collected during the planning portion of the NOAA Planning, Programming, Budgeting, and Execution System (PPBES). PPBES was instituted in 2004 to link NOAA's strategic vision with programmatic detail, budget development, and annual operating plans. For the 2007 assessment, representatives from each of the Goal Teams drew information from these plans and supplemented it, as needed, with direct communication with appropriate Goal Team program leads.

Information was collected and organized around three broad themes: (1) observation acquisition and transmission; (2) scientific data management; and (3) archive and access. Key acquisition and transmission issues include identifying observational requirements; monitoring of data streams to ensure accurate, complete, and secure data collections; and identifying, inventorying, and rescuing historical records from aging or non-digital media. Scientific data management includes calibration, validation, and data quality; data set description and documentation; adequate metadata; and the use of appropriate standard formats and protocols. Archive and access issues include bandwidth for transmission, storage and subsequent access capacities, automated and routine delivery of data to an appropriate archive center, and online catalogs and tools to aid in data discovery and use.

The assessment results can be seen in Figure 2. This chart presents the status of data management activities as reported by the NOAA Goal Teams. The responses were not normalized for differences in program responses; no quantitative value was assigned to "substantial" or "incremental" increases in resources, for example. A green cell indicates that the Goal Team is accomplishing 100 percent of the required activity or is able to do so with current resources. A yellow cell indicates the activity is partially accomplished or an incremental increase in resources is required to reach the 100 percent level. A red cell indicates the activity has not been accomplished or that substantial new resources are required. To aid in comparing the results to the 2005 assessment, symbols are added to indicate improvement (>) or decline (<) in capabilities since 2005.

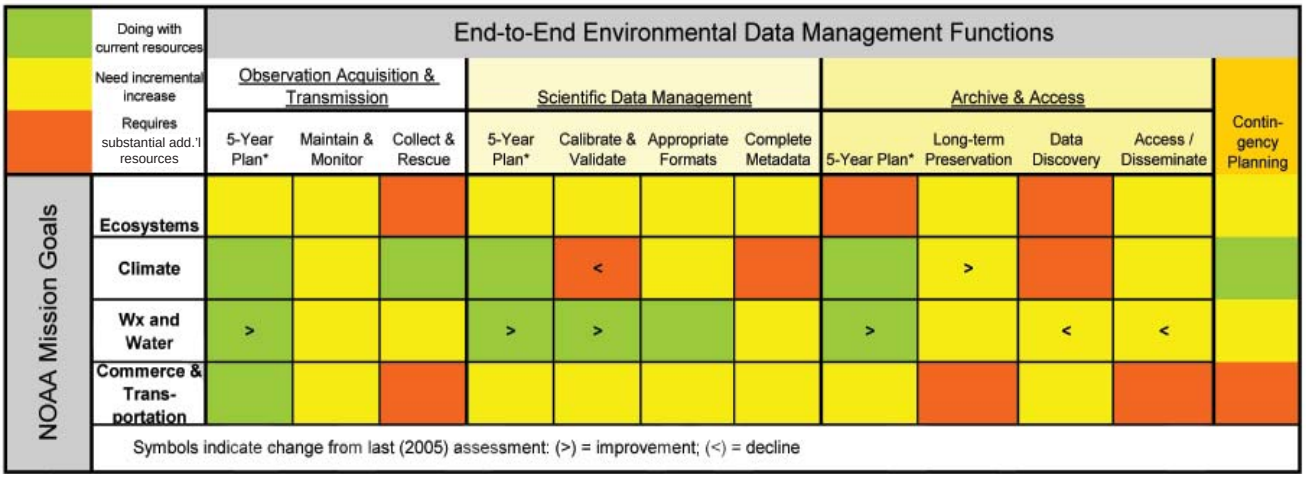


Figure 2. Assessment of NOAA data management capabilities by Mission Goal.

A fully successful Goal acquires high-quality data; analyzes, evaluates, documents, and preserves these data; and provides timely information and products to the end users of the services. These assessments, in combination with strategic planning on the enterprise level, provide NOAA with an integrated end-to-end approach for Earth environmental data and information management.

In addition to the summary chart in Figure 2, charts from each Goal Team, offering a program-by-program assessment, are provided in the Appendix. Compared to two years ago, the overall status of data and information management NOAA-wide shows some improvement. The Goal Teams are making incremental progress in some areas, while continuing to struggle with a number of common challenges. The most visible overall improvement is in the area of planning; several Goal Teams reported progress in generating their 5-Year Plans. Overall, data archive and access components remain the most challenging, with most Goal Teams reporting a need for substantially increased effort to meet requirements. These challenges are directly related to the ever-increasing growth of data streams requiring long-term stewardship and analysis from relatively newer data sources such as satellite sensors and numerical models for weather, climate, and ocean prediction.

Among the Goal Teams, Weather and Water programs report significant advances in the Scientific Data Management category, with the exception of metadata development. For example, the Tsunami Program has made substantial gains in developing plans, as well as in all data archive and access categories. The Climate Goal Team has made significant improvements in the quality of observations and is delivering data more efficiently, capturing higher quality observations, improving quality control and long-term preservation, and increasing spatial and temporal coverage of data collection. The most significant challenges for the Climate Goal are calibration and validation planning for the new series of the National Polar-orbiting Operational Environmental Satellite System (NPOESS) satellite sensors; data integration; adequate metadata; data access and dissemination; data discovery; rapidly growing data volumes; and data reprocessing for climate analysis to improve and expand climatological records. The Commerce and Transportation Goal Team reports some successes, including new, integrative data portals and upgrades to infrastructure to support continuity of data services. As with other teams, the Commerce and Transportation Goal Team continues to struggle with data archive and access functions, due in part to some data sets being in at-risk, non-digital formats, as well as to increased data volumes. Finally, the Ecosystem Goal Team reports little change from its 2005 status. Programs continue to maintain their capacity and incrementally modernize, though full success has not yet been achieved in any of the assessment categories.

The environmental data managed by NOAA activities advances or supports 6 of 10 National Essential Functions identified by the White House via the Homeland Security Council as the highest national priorities in the aftermath of major incidents of all origins. Thus, contingency planning for NOAA data systems is vital to ensure uninterrupted delivery of NOAA's products and services during disruptions to normal business. The Homeland Security Program Office (HSPO) of the Office of the Chief Information Officer leads NOAA-wide Continuity of Operations Planning (COOP) preparedness operations and participation in national level exercises. The Office of the CIO is also leading the development of enterprise security architecture to address full life-cycle data security requirements, and has identified IT infrastructure modernization initiatives to address data stewardship, processing, and dissemination.

Additionally, since 2005, and in partnership with Line and Staff Offices, the NOAA OCIO completed a comprehensive update of the Certifications and Accreditations (C&A) studies supporting 100 percent of NOAA's major systems. These C&A packages include information on disaster recovery and contingency planning.

Goal Team assessments of the status of their respective contingency planning status are included in Figure 2. COOP projects are underway in NOAA, especially within NNDC. In many instances, these Data Centers offer the vital mirrored or back-up site for other parts of NOAA. Much has been accomplished to ensure that NOAA will be able to continuously provide critical information to decision makers; however, Commerce and Transportation Goal data sets require incremental to substantial resources to fully implement contingency plans. Some historic and baseline data sets in jeopardy include water levels, ocean currents, and shoreline photogrammetry.

In viewing NOAA's data management challenges as a whole, the same five themes emerge from this assessment as from the 2005 assessment. These are:

- managing the increasing volume and diversity of data
- extending and filling gaps in environmental data records
- improving access to the long-term archive
- enabling integration of quality observations and products
- improving descriptions of data, metadata, formats, and processing steps

In the following paragraphs, we examine each of these challenges and, where appropriate, highlight recent successes or areas of continuing concern.

Successes and Challenges

Managing Increasing Data Volume and Diversity

NOAA data management systems face exponential growth in the volume of data (see Figure 3). In particular, new satellite sensing systems, including NPOESS and the NPOESS Preparatory Project (NPP), are expected to double incoming satellite-related data each year for the next several years. Additional large-array data streams from climate, weather, and ocean models are expanding rapidly, and newer multibeam and sidescan sonar systems are producing growing volumes of data from NOAA's hydrographic surveys. Yet another area of significant growth will be from weather radar. The National Weather Service Weather Surveillance Radar 88 Doppler—Next Generation Weather Radar (NEXRAD) data volume will double in the next 3 years and then increase again by an order of magnitude over the following 5 years with increased resolution and improved technology. Figure 3 illustrates growth in data volumes projected for the Comprehensive Large Array-data Stewardship System (CLASS) project, which houses NOAA's large-array data. Efforts are underway to build the infrastructure necessary to quality assure, describe, archive, and provide access to these data.

> The voyage of the British ship H.M.S. Challenger, from 1872-1876, is generally considered the first modern oceanographic expedition. Over roughly four years and 69,000 nautical miles, the Challenger collected data on the biology, chemistry, and physical properties of the world's oceans. The resulting data set ran to 29,500 pages and filled 50 volumes. Scanned and stored as digital image files (.pdf), the data might amount to a few billion bytes (a few gigabytes, or less if stored in a more compact format). By comparison, a single satellite can add almost 50 gigabytes of data per day to NOAA archives. Moreover, NOAA expects the current data volume from satellites to double annually over the next several years. In addition, the continued rapid evolution of computing technology will result in corresponding increases in the amount of output from weather, climate, and ocean models, which by some estimates will exceed the volume of satellite data. NOAA's data management technology must continue to evolve in parallel with its environmental sensing technology or face being overwhelmed.

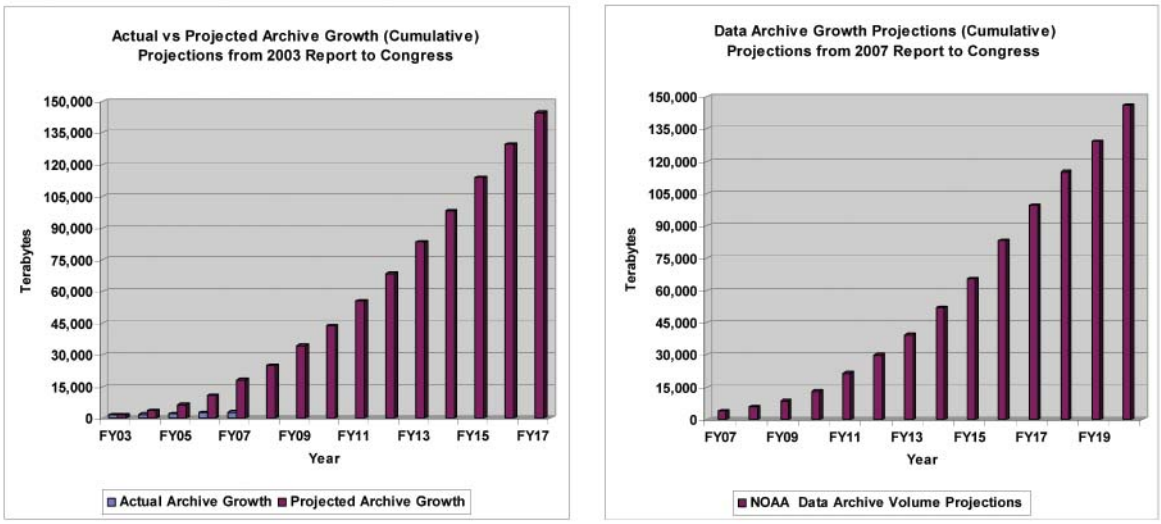


Figure 3. Growth in data volumes projected for the CLASS project, which houses NOAA's large-array data. Comparison of 2003 and 2007 projections reflect delays in several programs.

A major effort to address the data-volume gap is CLASS. CLASS is currently NOAA's premier online source for the storage of NOAA's Geostationary Operational Environmental Satellite (GOES) and Polar-orbiting Operational Environmental Satellite (POES) data. For these data, CLASS provides capabilities necessary to support long-term, secure storage of data, information, and metadata. It also enables access to these holdings through both human and machine-to-machine interfaces available through the NOAA National Data Centers and Centers of Data.

CLASS was originally conceived as a solution for a limited number of large-array (primarily satellite, weather radar, and weather model) data sets. However, under the current concept of operations, CLASS is envisioned as a true enterprise storage facility able to address data storage gaps across NOAA. For example, the Commerce and Transportation Goal Team reports that a substantial gap remains in the ability to handle the large and growing volume of hydrographic survey data that is growing due to the many recent advances in technology. Significantly larger data sets are being acquired in support of nautical charting during the performance of hydrographic surveying and shoreline delineation. Currently, large-volume, multibeam sonar data of up to 15 gigabytes per day per vessel and large airborne laser data sets with data rates up to 60 gigabytes per day are being acquired. Data handling, processing, and quality control issues are of major concern. Also, investigation is underway to determine the best methods to handle the digital side scan sonar and bottom backscatter data sets. These data are acquired in support of nautical charting efforts (up to 600 gigabytes of data per day per vessel), which are of interest to the Departments of Defense (DoD) and Homeland Security, and to geologists, many groups studying fisheries, and myriad other users. It is anticipated that these data, due to their volume and flux, will be stored and mirrored within CLASS.

Within this vision, CLASS will continue to focus on the information technology required to store and maintain the submitted data sets, for instance, by upgrading to new storage media as required and performing periodic data integrity tests. The NOAA Data Centers and Centers of Data will be responsible for providing CLASS with data and metadata, as well as maintaining access interfaces that are responsive to users. More information on CLASS and its future evolution is included in the Comprehensive Improvement Plan section of this report.

Extending and Filling Gaps in the Environmental Record

The establishment of trends of environmental change—their magnitude, direction, and geographic variation—generally requires the ability to detect small changes in the data against a noisy background (see Box). This is true whether the subject is atmospheric mean temperature, ocean currents, or marine ecosystems. Long data records that are quality-controlled and that use consistent or compatible measurement techniques and terminology are vital to detect patterns of environmental change. Programs such as the Climate Database Modernization Program (CDMP) support NOAA's mission to collect, assimilate, integrate, and effectively manage Earth observations on a global scale, ranging from atmospheric, weather, and climate observations to oceanic, coastal, and marine life observations. Many of NOAA's holdings, which are part of the U.S. National Archives, were originally recorded on paper, film, and other fragile media and stored at various NOAA Centers. Prior to CDMP, not only were these valuable data mostly unavailable to the scientific community, storage technology for the archives was not state-of-the-art. Without proper preservation of the media, the information they contained was in danger of being lost forever.

> Detecting climate change, understanding the associated shifts in specific climate processes, and evaluating the impacts of these changes upon human and natural systems requires a comprehensive set of consistent measurements taken over many decades. Detecting changes and variations in climate requires careful analysis of time series of sufficient length, consistency, and continuity to distinguish between the natural climate variability and any small, persistent climate changes. Confidently detecting the rate of climate change requires instrument accuracy and stability better than is generally required for weather research and most other scientific uses. Maintaining these measurements in an operational environment provides the best opportunity for sustaining the long-term data records needed to accurately assess climate variability and change. For the most part, metrics of instrument accuracies are inadequate for climate change detection; substantial investment is needed in scientific data stewardship where data are calibrated, quality-controlled, reprocessed, and assessed for validity. Scientific data stewardship budgetary needs are being assessed by NOAA's Climate Goal.

CDMP has greatly improved the preservation of and access to NOAA's holdings by migrating many of these resources to digital media. Digital images of many holdings are now available online, and millions of historical data records have been keyed and integrated into digital databases, as part of an ongoing process. In 2006, CDMP had over 60 separate

NOAA data rescue projects that spanned the full spectrum of NOAA activities, supporting all five line offices (see Figure 4). CDMP also works with U.S. Regional Climate Centers, State Climatologists, the U.S. Air Force, the World Meteorological Organization, and foreign meteorological services in Europe, Africa, Asia, and the Americas. These efforts benefit not only NOAA, but also researchers and data users across the globe. The increase in data accessibility and integrated global databases needed by today's climate and environmental data users validate the CDMP mission: to make major climate and environmental databases available via the World Wide Web.

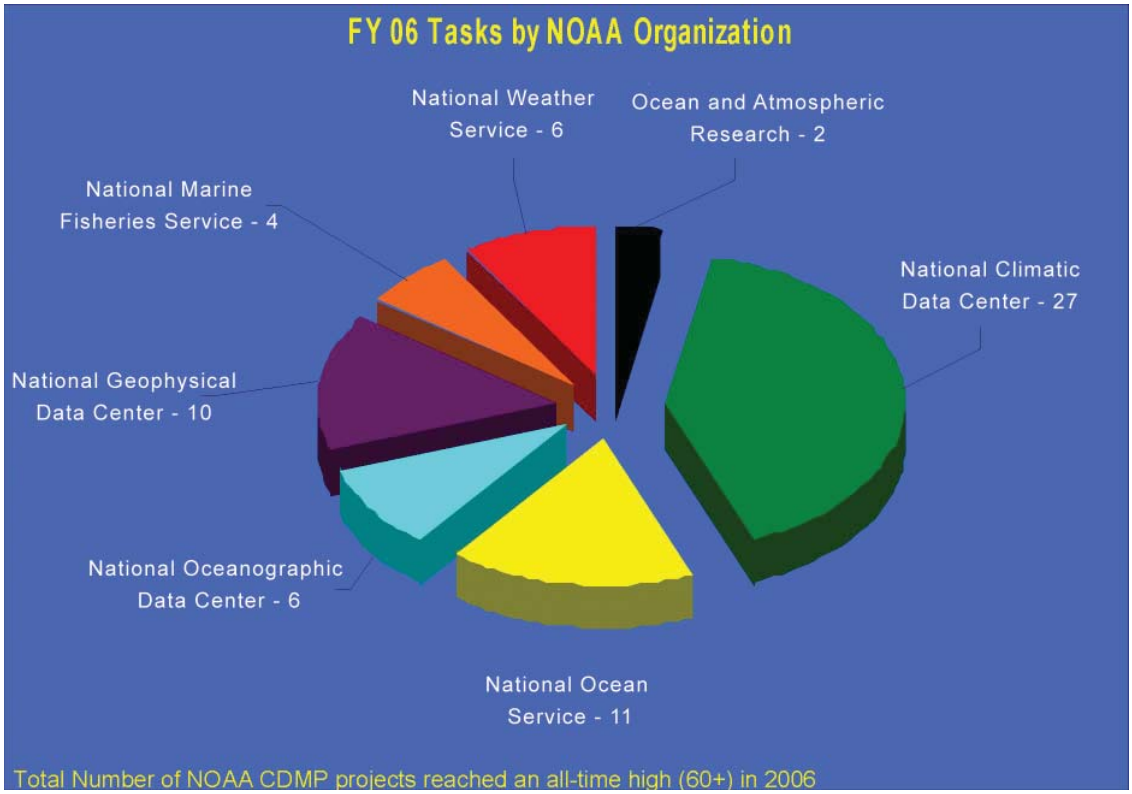


Figure 4. Breakdown of the CDMP projects by organization for FY 2006.

A particular challenge to extending the environmental data record, as reported by the Climate Goal Team, is the planned descoping of NPOESS. NPOESS, as originally configured, would have represented a significant step forward in the Nation's ability to deploy a comprehensive climate observing system. Many key climate variables would have been measured for decades. The original NPOESS was an enormous advance over the existing operational systems of POES and the Defense Meteorological Satellite Program (DMSP), and it sustained many of the capabilities demonstrated by the National Aeronautics and Space Administration (NASA) research missions. Unfortunately, the recent loss of climate sensors due to the NPOESS Nunn-McCurdy Certification places the overall climate program in jeopardy by creating, in effect, a gap in several key satellite data records. NASA and NOAA are seeking advice from the National Research Council (NRC) on the impacts of these climate sensor changes and are investigating what might be done to mitigate or recover the lost capabilities.

Improving Access to the Long-Term Archive

NOAA's vast and burgeoning data holdings will realize their full benefit to the public only if the data are accessible to a broad spectrum of user groups. NOAA continues to work toward expanding e-Government through a number of programs, such as the NOAA Virtual Data System (NVDS). NVDS is a Web-based system to support customer requirements for NOAA's global climatological, geophysical,

and oceanographic data from the three NESDIS Data Centers. The project began in the late 1990s and has continued to evolve in response to customer requirements.

Five components comprise NVDS: the NESDIS e-Commerce System (NeS), the NESDIS OnLine Store (OLS), the Climate Data Online (CDO) system, the Hierarchical Data Storage System (HDSS) Access System (HAS), and Geographic Information System (GIS) services. All Data Center orders that require payment are processed by NeS, with the OLS (integrated with NeS) handling online orders. CDO is the NVDS data model and system for delivering *in situ* climate data to customers. GIS services are fully compatible with Open Geospatial Consortium (OGC) standards supporting data discovery and visualization capabilities. HAS provides access to the HDSS tape robotics system (climate data archive) for radar, satellite, model, and other data. Over 1.5 petabytes[1] of data are now accessible online and 99 percent of all data orders are placed by online customers. Total online data delivery increased from less than one thousand gigabytes (or 1 terabyte) in FY 2000 to over 170 terabytes in FY 2006.

Since 2005, key NVDS improvements include:

- **Cross-system integration:** CLASS-hosted satellite data can now be ordered for off-line delivery on digital media via NVDS/NeS. For data stored on the HDSS tape robotics system, customers' orders via the National Operational Model Archive and Distribution System (NOMADS) (model data) and CLASS (satellite data) are supported by NVDS/HAS.
- **New services:** Numerous data sets and products have been added to NVDS and its GIS services. Several older systems have been retired as their capabilities were integrated into NVDS.
- **Infrastructure:** Hardware upgrades are allowing larger volumes of data requests to be processed, along with more complex queries and data mining activities.

The success of NOAA's e-Government activities has also been significant in that more data are now being downloaded by users each month while the rate of data ingest into the archives remains relatively steady. As depicted in Figure 5, access to NOAA's climate data has begun to show signs of an exponential increase in 2006. In past Reports to Congress, NOAA data management systems were described in terms of the amount of data flowing *into* the archives. Based on the trend of increasing demand for online climate data access, it is also important to address the amount of data being provided by the archive to ensure that users can access all the data they need and are not restricted by outgoing bandwidth.

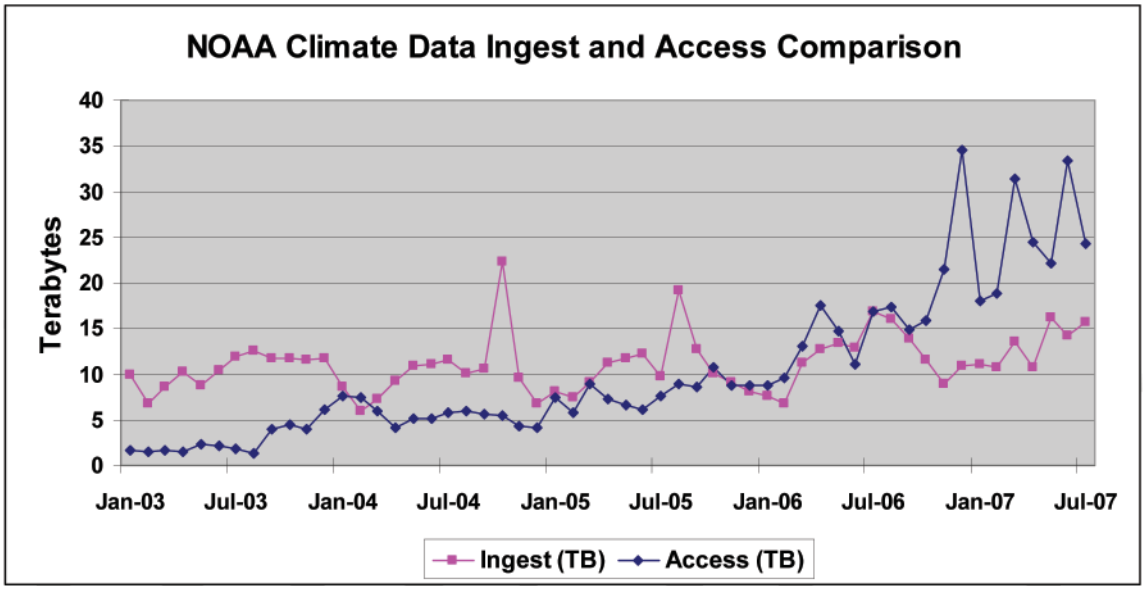


Figure 5. Growth in climate data online access (downloads in terabytes) relative to the amount of new data ingested. NOAA bandwidth for data access will need to grow accordingly.

[1] A petabyte is 10^{15} bytes or a million gigabytes.

NOAA is also working on methods to improve data accessibility through the use of advanced Web services, an emerging standard for electronic data exchange through the World Wide Web. For example, a demonstration site has recently been developed to offer access using this method to a small portion of NOAA's IOOS data holdings. The Portal for Oceanographic Services for IOOS Data and Information (POSIDIN) went online within the past year and has begun to provide access to near-real-time and historical data from the National Weather Service and National Ocean Service (go to http://opendap.co-ops.nos.noaa.gov/content for more information). Users may also access real-time listing of active observing stations and the status of their sensors. As the use of advanced web technology expands, it should offer indispensable data access to the wealth of NOAA's oceanographic data.

Enabling Integration of Quality Observations and Products

The ability to integrate data from multiple sources enhances all of NOAA's data intensive endeavors, from multidisciplinary, team-oriented research to the development of data products serving the Nation's aviation, transportation, and marine communities. Data integration builds upon standards for data and metadata and requires interoperability among independently operated data systems. NOAA is making progress toward interoperability, beginning with the development of GEO-IDE. GEO-IDE is a recently initiated effort to plan for data system interoperability across NOAA and beyond. It is a framework that enables the effective and efficient integration of NOAA's many existing systems and will serve as a guideline for future data system development efforts. The GEO-IDE effort will be discussed more fully in the Comprehensive Improvement Plan section of this report.

An example of the benefits of data integration is offered by the Meteorological Assimilation Data Ingest System (MADIS). MADIS is now being transitioned to operations within NOAA's Weather Service's Telecommunications Operations Center (TOC) in order to improve the utility of NOAA products and services and better serve the surface transportation enterprise, including roads, rails, transit, and pipeline/hazardous material operations. TOC has the requisite infrastructure to operate and maintain 24/7 operations with offsite backup.

MADIS will form the backbone of the National Surface Weather Observing System (NSWOS), providing the surface transportation sector (public, private, and academic/research organizations) with transportation-relevant, high-quality environmental data. The breadth of observations from public, private, and research sources available to NOAA requires a data management architecture that is able to accept this expanding suite of surface (and upper air) observations and their metadata. This architecture must also ensure the data are easily and effectively integrated to support surface weather and other NOAA and enterprise applications.

When MADIS has transitioned to operational status, disparate NOAA and non-NOAA observation networks will be available to reliably collect and integrate an increased volume of surface observations (tens of thousands of records). These disparate networks include, but are not limited to, the data systems listed in Table 1. MADIS will ensure NSWOS meets NOAA essential service requirements for activities, including dispersion modeling, homeland security, drought monitoring, and severe weather warnings and forecasts. This information, quality-controlled and in common formats, will have many uses, including supporting NOAA's essential services and creating models and GIS applications.

Observation System	Affiliation
Hydrometeorological Automated Data System	NOAA National Weather Service
Automated Surface Observing Systems	NOAA National Weather Service
Physical Oceanographic Real-Time System	NOAA National Ocean Service
National Water Level Observation Network	NOAA National Ocean Service
Continuously Operating Reference Stations	NOAA National Ocean Service
Climate Reference Network	NOAA National Environmental Satellite, Data, and Information Service
State Road Weather Information System	States
Networks	
FAA and State Automated Weather Observing Systems	Federal Aviation Administration
U.S. Forest Service Remote Automated Weather Stations	U.S. Forest Service
USDA SNOTEL	U.S. Department of Agriculture
Federal, State, and Local Stream and River Gauges Cooperative Observer Observation Program Network	NOAA National Weather Service

Table 1. Planned Data Contributors to NOAA MADIS.

A second example highlighting recent successes in data integration is NOAA's Tsunami Program. To support the national strategy for minimizing the impact of tsunamis and meet the requirements of the Tsunami Warning and Education Act, NOAA relies on a network of global data, acquired and processed in real time, as well as high quality, retrospective global databases supporting advanced scientific modeling. These data include seismic data, Deep-ocean Assessment and Reporting of Tsunamis (DART™) buoy data, coastal water level (tide gauge) data, nearshore and deep ocean bathymetric data, and historical tsunami event and inundation data (see Figure 6). At the time of the last Report to Congress, NOAA had just implemented a new Tsunami Program and was facing significant challenges in its observing systems, the acquisition and dissemination of real-time data, and the long-term stewardship of these data to enable research and risk assessment. In the two years since, the Tsunami Program data management improved, eliminating "red" cells in observation acquisition and transmission, scientific data management, and archive and access. NOAA has upgraded and expanded its seismic, DART™, and coastal observing systems; is replacing obsolete transmission systems with more capable and reliable systems; and is upgrading and expanding its data stewardship capabilities.

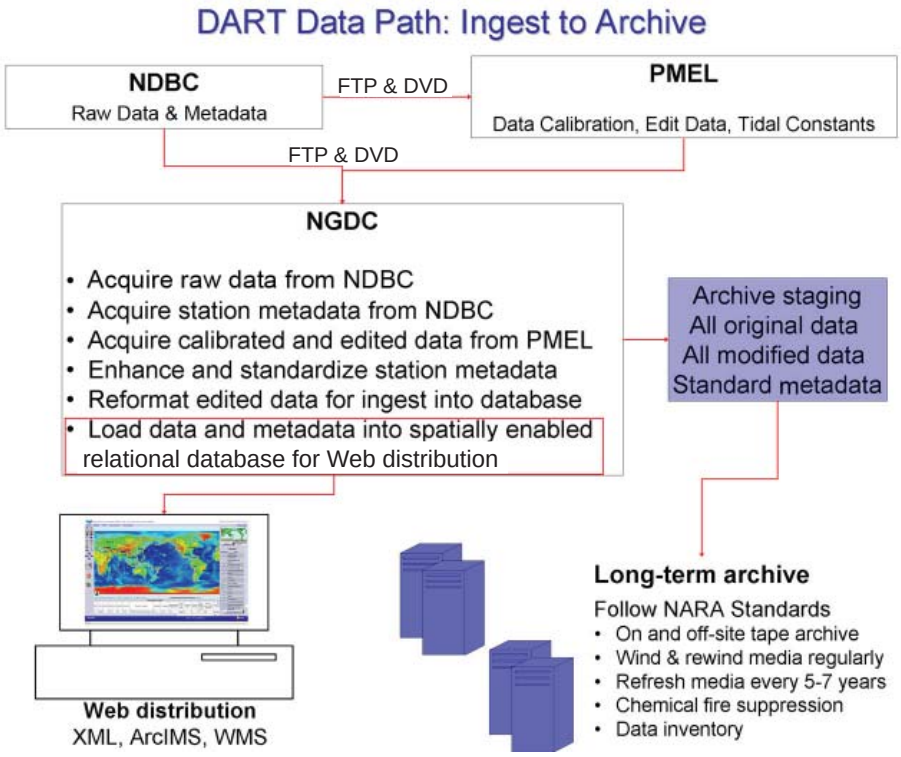


Figure 6. The Tsunami Program integrates data from across NOAA to create products that save lives.

The tsunami-related data archive has grown from 5 gigabytes to over 1,700 gigabytes of data, all archived to National Archive and Records Administration standards with standards-compliant metadata available online through Web Map and Web Feature or Web Coverage Services, supporting both the GEOSS GEO portal and the modeling, mapping, and assessment activities required to minimize the effect of tsunamis. In addition, the Tsunami Program completed a first planning document identifying the flow of data from acquisition to archive, documenting observation transmission formats and pathways, single points of failure and vulnerabilities, distribution of tsunami products, and eventual long-term preservation and access of data essential for the Nation's Tsunami Program. Based on data available from the archive and in partnership with the U.S. Geological Survey and the National Tsunami Hazard Mitigation Program, the Tsunami Program also completed the first assessment of tsunami hazard for all U.S. coasts. Over the next year, NOAA's high-resolution coastal tide data will be added to the long-term archive and will be made available, following the same standards-based protocols, and work will begin on the archive and access to tsunami models and model outputs.

A final set of examples points the way forward for the integration of NOAA's ecosystems data. NOAA is embarking on the development of Integrated Ecosystem Assessments (IEA) as a critical element of the agency's commitment to Ecosystem-based Approaches to Management (EAM). The IEA provides quantitative analysis, assessments, and ecological forecasts at the regional ecosystem scale. (NOAA has identified seven regional ecosystems in the U.S. Exclusive Economic Zone: the Northeast Shelf, Southeast Atlantic, Caribbean Sea, Gulf of Mexico, California Current, Alaska Complex, and Pacific Islands.) The IEA focuses on integrating existing datasets and research to build models for ecological forecasts, conduct assessments, and develop decision support tools, as opposed to collecting new observations or conducting new research. The data and information management challenges involved with implementing IEAs are substantial. The IEA requires integrating traditionally disparate data sets. For example, forecasting the effects of climate changes on ecosystems requires combining biological data with physical data.

Two specific projects in the Gulf of Mexico are building on existing infrastructure to improve and enhance the dissemination of and access to ecosystem data. The first project is the Northern Gulf of Mexico Data Services Pilot. This pilot project serves to meet the Subcommittee on Integrated Management of Ocean Resources' (SIMOR) objectives of assisting in the coordination and implementation of Internet-based data access and delivery services tailored for the needs of State-level coastal and marine resource managers by drawing on existing Federal and State data and information sources. Federal agencies, including NOAA, NASA, the Environmental Protection Agency (EPA) Gulf of Mexico Program, Naval Research Laboratory-Stennis (NRL), and United States Geological Survey (USGS), working with State and academic partners in the Gulf, are providing access to ecosystem metadata, data, and data services through a common portal, enabling discovery and access to data through a semantic data catalog, and geospatially enabling the data for Internet display.

The second project is the EPA-NOAA Harmful Algal Blooms Observing System (HABSOS) website. It is a regional, Web-based data and information dissemination tool for the Gulf of Mexico developed as a regional coalition of U.S. and Mexican federal and state agencies. HABSOS and the more recent Bi-National HABSOS were developed and continue to be supported by the EPA Office of Research and Development, the EPA Gulf of Mexico Program, and the NOAA National Coastal Data Development Center. The HABSOS website provides a secure data entry tool (available in both English- and Spanish-language versions) for collection of cell count observations of the algal species *Karenia brevis*. Data entered into the system are aggregated and available for display in the HABSOS Internet Mapping Service (IMS). Online assessment and analysis of HAB events are enhanced through the integration of *in situ* observations, surface forecasts, and powerful satellite imagery products into the IMS. A Web link is also provided to the official NOAA HAB Bulletin.

Improving Descriptions of Data, Metadata, Formats, and Processing Steps

NOAA invests substantially in the acquisition of quality data, resulting in a national asset that must be protected for current and future generations. Maintaining a digital data archive means more than simply storing files on an appropriate digital medium. The data must remain usable over time, which means, among other things, that the data must be described thoroughly and accurately.

The production of these data descriptions, or metadata, is costly and time-consuming. For this assessment, no Goal Team indicated that there is current, complete, and authoritative metadata in appropriate formats describing all the Goal's data and products. Most Goal programs report incremental progress in achieving this goal, though the Climate Goal reported—as in 2005—that progress toward developing complete metadata necessary for determining long-term climate trends and predictions and for mitigation strategies has been limited by fiscal constraints.

This 2007 assessment shows that NOAA has continued over the past two years to make progress toward a comprehensive data management architecture that safeguards NOAA's priceless data assets. In particular, progress has been made in developing data management plans that address the ongoing challenges of increasing data volume, extending the environmental data record; improving access to data; enabling data integration across NOAA; and improving the state of NOAA's descriptive metadata. In addition (as described above), NOAA is continuing to implement existing programs and is initiating new programs to address these challenges. However, it is clear that these challenges will likely remain for some time.

III. Comprehensive Improvement Plan

Public Law 102-567, Sec 106 (c) (2): "The report shall -

(A) set forth modernization and improvement objectives for the 10-year period beginning with the year in which the plan is submitted, including facility requirements and critical new technological components that would be necessary to meet the objectives set forth;

(B) propose specific agency programs and activities for implementing the plan;

(C) identify the data and information management, archival, and distribution responsibilities of the National Oceanic and Atmospheric Administration with respect to other federal departments and agencies and international organizations, including the role of the National Oceanic and Atmospheric Administration with respect to large data systems like the Earth Observing System Data and Information System; and

(D) provide an implementation schedule and estimate funding levels necessary to achieve modernization and improvement objectives."

Historically, the development of NOAA's data management systems was focused solely on a particular project, discipline, or observing system capability. While these legacy systems were individually efficient, many of the current and potential uses of NOAA data and services are interdisciplinary in nature. This requires access to a wide range of NOAA's capabilities, and the disparate user interfaces, data models, protocols, formats, etc., of the legacy systems present a barrier to their effective access and use. The need for more common approaches to the development of data systems and services built upon data and information standards has been apparent for some time.

(A) NOAA's Data Management Plan

Plans for the development of an integrated NOAA data system capability have moved forward significantly since 2005 through the work of the Data Management Integration Team (DMIT), a group comprising NOAA data management specialists with representation from all NOAA line offices, goal areas, and the CIO's office. In 2005, the NOSC and its DMC directed the DMIT to develop GEO-IDE. In early 2007, NOAA created and filled the position of the Data Management Information Architect to chair DMIT and lead the GEO-IDE initiative.

Through a coordinated, NOAA-wide effort, DMIT has developed a GEO-IDE Concept of Operations (CONOPS) that outlines the scope, objectives, and approach for this initiative. The GEO-IDE vision is a framework that enables the effective and efficient integration of NOAA's many existing systems and will serve as a guideline for future data system development efforts. It proposes the use of a services-oriented architecture and a standards-based data and information infrastructure that will provide access to the full range of underlying data system capabilities. It also outlines a process for NOAA to adopt or adapt data and information system standards that will advance the GEO-IDE objectives. In addition to the CONOPS, DMIT has published a draft GEO-IDE Implementation Plan and a Standards White Paper with an initial list of recommended standards.

GEO-IDE planning and coordination will continue to be the responsibility of DMIT and a small project management office. They will define and implement the processes required to establish NOAA-wide data management standards and guidelines, evaluate new and emerging data management technologies, and sponsor or oversee the development of infrastructure components. Through GEO-IDE, NOAA will be able to identify, endorse, or develop standards and protocols to effectively migrate legacy systems toward a common vision. GEO-IDE also defines and prioritizes specific actions to pursue and proposes organizational responsibilities to implement integrated data management capabilities. The true implementation of GEO-IDE will be a long-term effort that will be accomplished when all NOAA's data system development projects are fully compliant with the GEO-IDE principles, standards, and guidelines.

The objectives and approach of GEO-IDE are fully consistent with and support the Strategic Plan for the U.S. IEOS, which in turn is the U.S. contribution to GEOSS. All three complementary initiatives

are attempting to remove barriers that prevent or inhibit the most productive use of Earth observation data applied to a wide range of societal benefit areas. Where GEO-IDE is focused on improving access to NOAA's data and information systems and services, USGEO and GEO are addressing the same challenges faced across Federal agencies and the nations of the world. In each case, the initiatives are leveraging the technical innovations and best practices of their members to develop common, standards-based components and an interoperable set of data and information services. GEO-IDE provides NOAA with a mechanism to both contribute to and benefit from the efforts underway at the national and international level (see Figure 7).

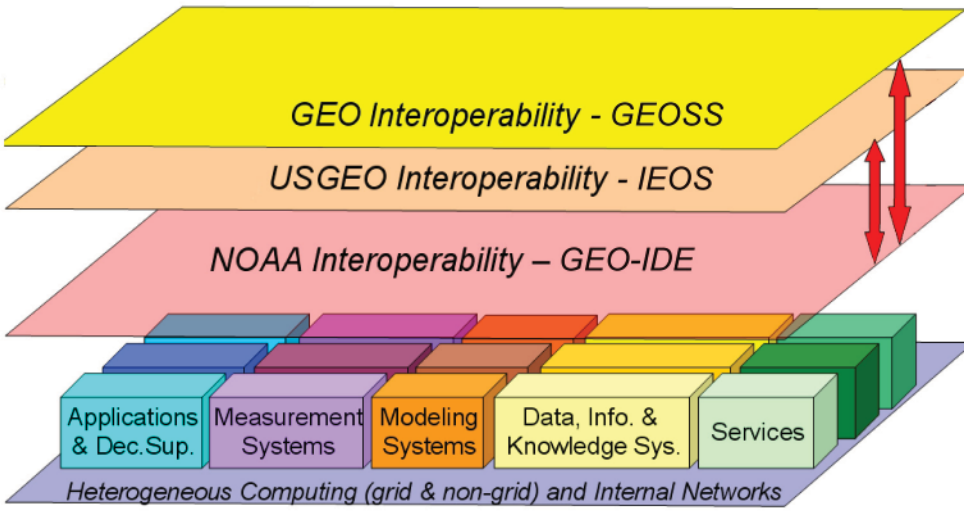


Figure 7. Elements of NOAA's plan for interoperability and linkage to USGEO and GEOSS.

(B) Implementing the Plan

Development of the GEO-IDE vision is ongoing. However, NOAA is taking concrete steps to provide integrated data products and services. Two examples of data management programs that are meant to embody the GEO-IDE vision are the Scientific Data Stewardship (SDS) program and CLASS.

Scientific Data Stewardship Program

The SDS program's goal is the production of high quality Climate Data Records (CDRs) for data from the atmosphere, oceans, and land surface, where the data are identified as essential climate variables within the Global Climate Observing System (GCOS). The SDS program seeks to support the regular, operational production of CDRs for data from the atmosphere, oceans, and land surface. This production requires collaboration between experts in the climate community and experts in data management. It also must be informed by scientific application and associated user feedback on the accessibility and usability of the produced CDRs.

SDS has identified three development areas for operational and regular production of high-quality CDRs:

- **Operational quality assurance of ingested data and regular monitoring of data quality and provenance during production:** For example, by identifying and using appropriate tools for detecting and diagnosing subtle spatial and temporal biases in the incoming data stream and in the software that transforms that stream into other data products.
- **Generation of authoritative, long-term records:** Through rigorous data analysis and research that

will validate and improve these records, particularly by intercomparing multiple data sources and reprocessing entire data streams. The authoritative nature and vitality of CDRs will be maintained through peer reviews, user recommendations, and independent processing of data that produces alternative versions of essential climate variables.

- **Configuration management for CDRs:** To ensure data usability now and in the future by providing documentation of data itself, as well as of the archival process, in a way that preserves the context that will allow future users to understand, modify, and use the data.

The SDS program seeks to encompass the full range of institutional diversity within the climate community. To that end, this program seeks approaches that allow organizations producing and archiving CDRs to maintain local autonomy within a context that encourages responsible participation in federations that foster increased data sharing, interdisciplinary data understanding, and improved assessments of data quality.

The Comprehensive Large Array-data Stewardship System

The Assessment section of this report described the evolution of CLASS over the past two years, from a system geared primarily toward end-to-end management of a few satellite data sets to a true "One-NOAA" data-archive storage system. Within the GEO-IDE framework, CLASS will continue to evolve as a major component in NOAA's data management architecture.

CLASS will be an open system architecture whose design will incorporate standards and processes implemented under GEO-IDE. This approach will accommodate interoperability within existing and new data systems as they also adopt interoperability standards. CLASS will:

- Provide an enterprise data storage capability that centralizes NOAA's environmental data access
- Provide access interfaces to support data analysis, including reprocessing, reanalysis, etc., via the NOAA National Data Centers and Centers of Data
- Provide a cost-effective architecture that is adaptable and expandable to handle many types of data
- Optimize the use of existing and preferred infrastructure capabilities with new investments
- Provide a tracking system that allows for the tracking of all data sets from originator-provided data to any subsequent versions of the data

The development of CLASS is expected to be a long-term, evolutionary process, as current and new "campaigns" (sources of data) are incorporated into the CLASS architecture and as the GEO-IDE vision continues to take shape.

(C) NOAA's Data Management Responsibilities

The 2005 Environmental Data Management Report to Congress detailed NOAA's statutory responsibilities for data management and its many relationships with other Federal and international agencies and efforts. These responsibilities have not changed; the NNDCs have the unique responsibility for the long-term management and stewardship of the bulk of NOAA's data in conformance with the regulations and procedures of the National Archives and Records Administration, and NOAA continues to take responsibility for non-NOAA national environmental data sets through agreements with other Federal agencies.

In 2006, at the request of NOAA, the NRC empanelled a committee to provide advice on archiving and access to the broad range of environmental and geospatial data collected by NOAA and its partners. With limited resources and enormous growth in data volumes, NOAA was seeking input on how to identify the observations, data, and derived products that should be preserved in perpetuity and made readily accessible versus those that require limited access and storage lifetimes. At the end of 2006, the

NRC committee produced a preliminary list of principles and guidelines on archiving. The NRC committee is now refining and expanding this preliminary set of principles and guidelines for data archiving into a final report that will also address the extent to which a wide variety of data sets should be made accessible. This final report, expected in Fall 2007, will provide NOAA with advice on what types of data should be archived indefinitely (for at least 75 years); what types of data could be stored for shorter durations under budgetary constraints; and how best to provide access to different variables, data sets, and derived products.

A unifying context for defining data management responsibilities across NOAA, the Federal government, and internationally is provided by the USGEO nationally and GEO internationally. NOAA has taken steps to begin implementation of its portion of the national IOOS[2], which will benefit NOAA's ocean and coastal programs and their customers, and start building an integrated data management infrastructure applicable to all of NOAA's environmental data and products.

The Integrated Ocean Observing System

The nationally scoped IOOS[3], as called for in the President's U.S. Ocean Action Plan[4], is a system-of-systems concept that leverages the Nation's existing capacity in ocean observations, modeling, and analysis—coupled via a Data Management and Communications (DMAC) component. It is the ocean portion of the broader national IEOS and the U.S. contribution to related international efforts—the Global Ocean Observing System and GEOSS (see Figure 8).

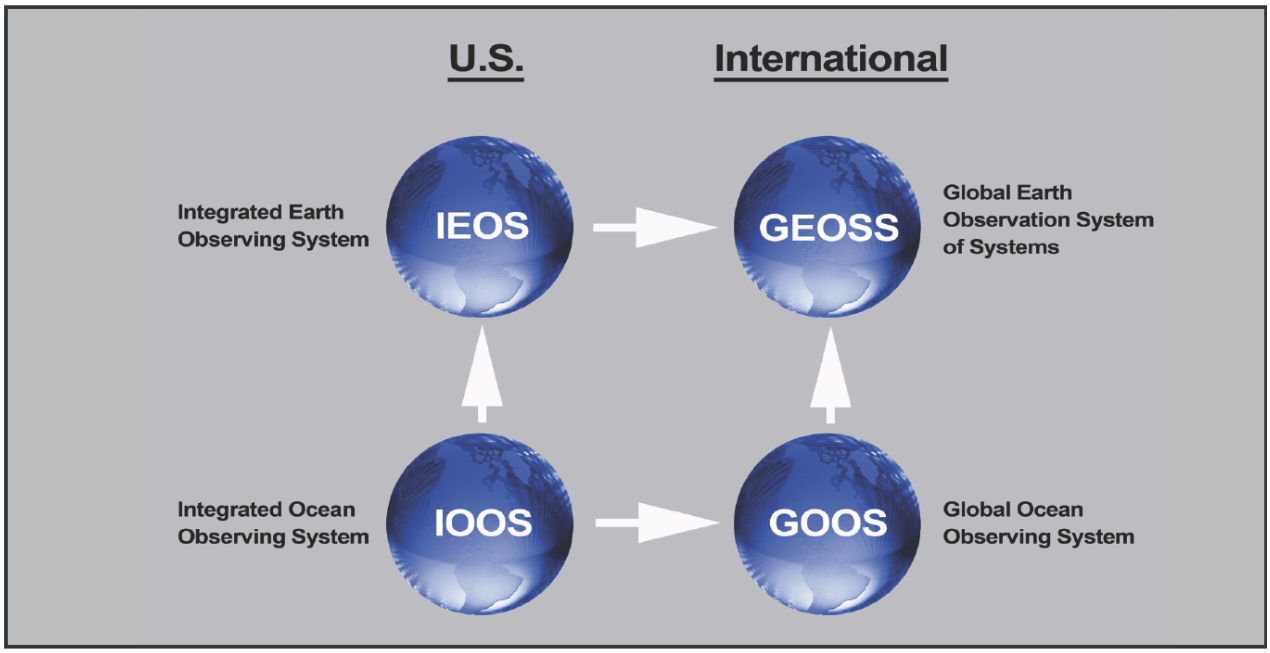


Figure 8. Relationships between elements of the Global Earth Observation System of Systems.

Implementation of the DMAC component of IOOS will have significant benefit for NOAA's overall data management capability. Its concept and operating principles align directly to the broader programs cited above and address the end-to-end perspective necessary for a robust and extensible data management framework. Its implementation will provide valuable lessons-learned and reusable components for implementing the NOAA GEO-IDE.

[2] Footnote to FY08 President's budget request for IOOS.
[3] Footnote to IOOS Development Plan.
[4] Footnote reference to Ocean Action Plan and IOOS as part of GEOSS page.

With the FY 2008 budget request, NOAA proposes building an Initial Operating Capability for its portion of IOOS that includes development of a data integration framework for a limited subset of five ocean variables: temperature, salinity, sea level, surface currents, and ocean color. The initial framework will implement a subset of DMAC functionalities to support a diverse set of four NOAA products or activities: IEAs, harmful algal bloom prediction, coastal inundation modeling, and hurricane intensity modeling. The effort also includes systematic test and evaluation of resulting product enhancements and benchmarking of new performance specifications for a decision to proceed to operational use. The initial capability also will be assessed for its performance with respect to the NOAA Target Architecture and GEO-IDE objectives. Success with this initial capability will lead to expansion of the framework to encompass additional ocean variables and data management functionalities.

(D) Implementation Schedule

NOAA identified several projects to modernize data management, ranging from real-time acquisition, processing, distribution, and modeling to description, archive, and access, and reprocessing and rescue. These modernization projects are a critical part of the total NOAA data management effort. Figure 9 shows the major milestones and implementation dates for the modernization projects identified in the accompanying budget table (see Figure 10). These capabilities support NOAA's development of a comprehensive, coordinated, and sustained Earth observation system to collect, process, disseminate, and archive improved data, information, and models.

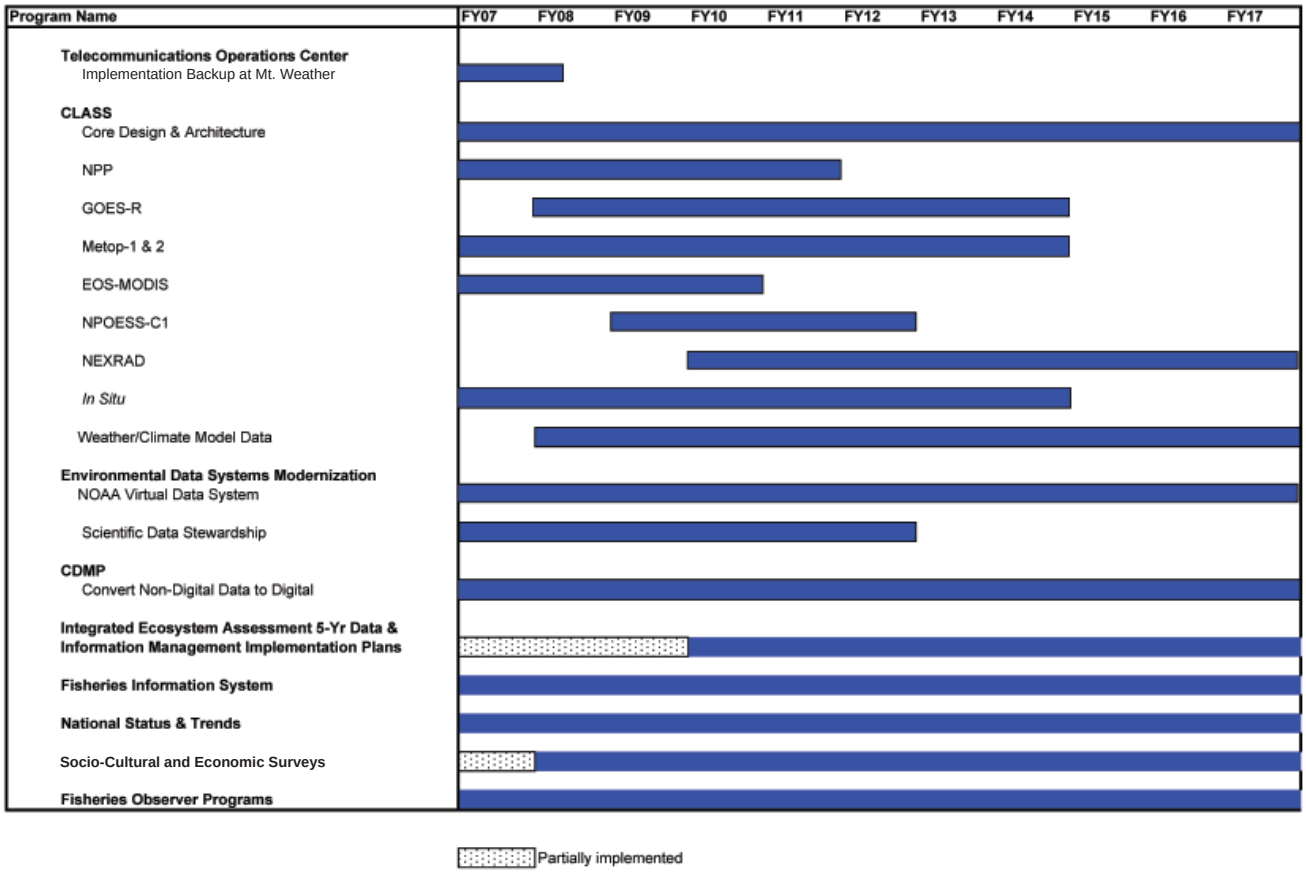


Figure 9. Implementation schedule for NOAA data and information modernization activities.

	FY06 Enacted ($K)	FY07 President's Budget ($K)	FY07 Enacted ($K)	Modernization Activity
Telecommunications Operations Center (TOC)	3502	6007	6000	TOC backup
Comprehensive Large Array Data Stewardship	8876	6476	7011	Data archive volume
Environmental Data System Enhancement	9256	9346	9296	Data interoperability, metadata
Climate Data Base Modernization	20810	4063	16526	Data rescue
Integrated Ecosystem Assessment 5-Yr Data & Information Management Implementation Plans	910.5	999	999	Data access, dissemination, archive
Fisheries Information System	37400	37770	40000	Operations, enhancement, development
National Status & Trends	125	125	125	Continuity of operations
Socio-Cultural and Economic Surveys	1825	13041	3201	Continuity of operations
Fisheries Observer Programs	805	1653	1653	Data acquisition, access, dissemination, & archive

Figure 10. Funding for NOAA data modernization.

Telecommunications Operations Center (TOC)

The first step in this process is rapid, reliable acquisition, processing, and distribution of real-time data and products. The ongoing improvements to NOAA's TOC help address these challenges. TOC is NOAA's 24/7 nerve center for real-time environmental data dissemination, monitoring the Global Telecommunication System (GTS), Advanced Weather Interactive Processing System (AWIPS), Wide Area Forecast System, NOAA Weather Wire Service, and NOAA Weather Radio status. In addition to these national and international circuits, TOC also maintains connectivity with DoD, the Federal Aviation Administration, and the Department of Homeland Security. During FY 2006-2007, TOC completed the upgrade to message switch capabilities and implemented a backup facility at Mount Weather, Virginia, to ensure uninterrupted operations and manage the increased flow of data. TOC replacement is designed to meet future data volume requirements, with expansion capacity to 500 percent of the current average traffic volume.

CLASS

As noted earlier, under the current concept of operations, CLASS is seen as a true enterprise storage facility able to address the data storage gaps existing across NOAA. Within this vision, CLASS will continue to focus on the information technology required to store and maintain the submitted data sets—for instance, by upgrading to new storage media as required and performing periodic data integrity tests. CLASS continues to support the efforts under development in the Environmental Data Systems Modernization project.

Environmental Data Systems Modernization

The Environmental Data Systems Modernization project planned for FY 2006-2015 focuses on developing an integrated, virtual NOAA data system and improving the scientific stewardship of NOAA data. NOAA's virtual data system, of which CLASS is a key component for the large-array data, will provide improved seamless access to data, metadata, and products. This is a first step in fulfilling the U.S. commitment to disseminate improved data, information, and models in support of USGEO and GEO. The Scientific Data Stewardship Program, a new NOAA initiative in FY 2006, will provide high-quality climatology of the atmosphere, oceans, and Sun-Earth geophysical environment. Tasks include monitoring observing systems, generating near-real-time climate records, reprocessing, and generating new products—made possible by extending the data record through reprocessed or rescued data.

Climate Database Modernization Program (CDMP)

CDMP supports NOAA's mission to collect, integrate, assimilate, and effectively manage Earth's environmental observations on a global scale. Observations collected from various platforms and sensors include the whole biosphere from the oceans to the upper atmosphere. They also extend to processing and archiving NOAA's satellite observations and data. CDMP has greatly improved the preservation and access to these holdings by migrating many of the resources (e.g., paper, microform, publications) to new digital media and the Internet. Digital images in user-friendly formats are now available online; millions of historic data records have been keyed and integrated into various NOAA databases. As of 2007, CDMP has extended the historic climate record back nearly 100 years. During the next decade or so, CDMP will continue to acquire relevant historic environmental data and information, converting around 80-110 million records per year and making these data easily accessible to the user community.

Integrated Ecosystem Assessments (IEA)

The IEA integrates existing data to support quantitative analysis, assessments, and ecological forecasts at the regional ecosystem scale. The first IEA is being developed in the California Current regional ecosystem through a multiagency and university collaboration called the Pacific Coast Ocean Observing System (PaCOOS). The present focus of PaCOOS is on access to biological and physical data at existing NOAA sites. In 2008, PaCOOS will start to develop web-based tools to enable integration of biological and physical data into products, including an IEA.

Fisheries Information System (FIS)

FIS helps to improve the collection, management, and dissemination of commercial and recreational fisheries statistics. FIS is a nationwide framework planned cooperatively by Federal and State agency partners to expand, adapt, and integrate data collections to meet emerging national and regional information needs. Planned or underway improvements include: (1) an improved national permits system, which provides one-stop shopping for permit customers, harmonizes permit processes, allows sharing of permits information both within and among regions, and creates a national registry of permit holders; (2) a fishery information metadata system, which documents national fishery information assets and procedures; (3) a national electronic reporting initiative to improve both the delivery and management of fisheries data; and (4) initiatives to develop national and regional one-stop shopping access to well-integrated commercial and recreational fisheries information through a national system of systems.

NOAA's National Marine Fisheries Service (NMFS) is now required by the Magnuson-Stevens Reauthorization Act to establish and implement a regionally based registry program for recreational fishermen and a new data collection program that improves the quality and accuracy of recreational information. NMFS initiated work with the interstate marine fisheries commissions, State agencies, regional fishery management councils, and constituents in 2007 to identify, test, and implement improved methods for survey sampling, statistical estimation, and information management. The new Marine Recreational Information Initiative (MRII) will improve the accuracy, statistical precision, and timeliness of recreational fishery catch monitoring surveys and will speed integration of State and Federal fisheries information into regional and national networks, thereby increasing its utility and availability for stock assessment scientists, fishery managers, and the public.

National Status and Trends (NS&T) "Mussel Watch"

Mussel Watch, a component of the National Status and Trends Program, is the Nation's longest running coastal contaminant monitoring program. The goal of the Mussel Watch Program is to describe the spatial distribution and temporal trends in coastal toxic contamination in areas not in close vicinity of industrial or sewage outfalls or toxic hot spots, thus allowing for assessment of historical trends of coastal contaminants over time. The data are used by the scientific and management communities, agency personnel, and the general public to assess water quality, and environmental and human health impacts,

and to plan resource management activities. The Mussel Watch Program samples over 300 sites on all of the Nation's coasts, including the Great Lakes; data have been gathered for 21 continuous years. The Mussel Watch Program is continuing its long-term monitoring while currently being the first program to explore the national distribution of a new class of flame retardants contaminants, the polybrominated diphenyl ethers.

National Fisheries Observer Program

The National Fisheries Observer Program run by NOAA's NMFS provides information on the catch and bycatch of managed species and the bycatch of nontarget and protected species in U.S.-based fisheries. Ongoing activities within the program include the National Bycatch Report, the Data Accessibility Standards Project, the Alaska Interactive Data Access System, the Pacific Islands Longline Observer Data System, the Southwest Region Data System for Highly Migratory Species, and the Southwest Electronic Reporting System.

IV. Conclusion

NOAA's Goal Teams completed a self-assessment that allows comparison with a similar assessment finished in 2005 of the status of data management across NOAA. NOAA continues to face the same overall data management challenges it faced two years ago: rapidly increasing data volume and diversity; extending and filling gaps in environmental records; improving access to the long-term archive; enabling data integration; and improving descriptions of data, formats, and processing steps. However, within each NOAA Goal Team incremental progress is being made and NOAA can point to real successes in addressing each of these ongoing challenges.

NOAA has made substantial progress in planning for an integrated, "One-NOAA" data management architecture. The cornerstone of this planning effort is the GEO-IDE, which lays out a framework that enables the effective and efficient integration of NOAA's many existing systems and will serve as a guideline for future data system development efforts. Development of the GEO-IDE vision will be directed by NOAA's newly appointed Data Management Information Architect.

NOAA has reached out to coordinate with other agencies and with other countries to develop environmental observation and data management strategies. GEO-IDE provides a mechanism to coordinate the data management contributions of NOAA to USGEO and GEO and to leverage the approaches and advances of IEOS and GEOSS within NOAA data system initiatives. With the FY 2008 budget request, NOAA proposes building an Initial Operating Capability for its portion of IOOS that includes development of a data integration framework for a limited subset of ocean variables. Finally, NOAA has received the National Research Council's report entitled "Environmental Data Management at NOAA—Archiving, Stewardship, and Access." NOAA concurs with the recommendations and is assessing how current practices may be augmented by the principles in the report.

Abbreviations and Acronyms List

AWIPS	Advanced Weather Interactive Processing System
CDMP	Climate Database Modernization Program
CLASS	Comprehensive Large Array-data Stewardship System
Dec Sup	Decision Support
DMAC	Data Management and Communications
DMC	Data Management Committee
DMIT	Data Management Integration Team
DMSP	Defense Meteorological Satellite Program
EOS	Earth Observing System
FY	Fiscal Year
GCOS	Global Climate Observing System
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
GOES	Geostationary Operational Environmental Satellite
GOOS	Global Ocean Observing System
IEOS	Integrated Earth Observation System
IHO	International Hydrographic Organization
IOC	Intergovernmental Oceanographic Commission
IOOS	Integrated Ocean Observing System
IT	Information Technology
IUOS	Integrated Upper-Air Observing System
MODIS	Moderate Resolution Imaging Spectroradiometer
MOU	Memorandum of Understanding
NARA	National Archives and Records Administration
NASA	National Aeronautics and Space Administration

NCDC	National Climatic Data Center
NDBC	NOAA's National Data Buoy Center
NDE	NPOESS Data Exploitation
NEXRAD	Next Generation Weather Radar
NGDC	National Geophysical Data Center
NOAA	National Oceanic and Atmospheric Administration
NODC	National Oceanographic Data Center
NOSC	NOAA Observing Systems Council
NPOESS	National Polar-orbiting Operational Environmental Satellite System
NPP	NPOESS Preparatory Project
NWS	National Weather Service
PaCOOS	Pacific Coast Ocean Observing System
PMEL	Pacific Marine Environmental Laboratory
PPBES	Planning, Programming, Budgeting, and Execution System
POES	Polar-orbiting Operational Environmental Satellite
TB	Terabyte (one trillion [10^{12}] bytes)
TOC	Telecommunications Operations Center
USGEO	United States Group on Earth Observations
WDC	World Data Center
WMO	World Meteorological Organization
Wx	Weather

Appendix

Goal Team Assessment Summaries

Ecosystem Goal by Program

Goal Team Program	End-to-End Environmental Data Management Functions											
	Observation Acquisition & Transmission			Scientific Data Management				Archive & Access				Contin-gency Planning
	5-Year Plan*	Maintain & Monitor	Collect & Rescue	5-Year Plan*	Calibrate & Validate	Appropriate Formats	Complete Metadata	5-Year Plan*	Long-term Preservation	Data Discovery	Access / Disseminate	
Habitat												
Corals									<			
Coastal and Marine Resources												
Protected Species												
Fisheries Management												
Aquaculture												
Enforcement												
Ecosystem Observations												
Ecosystem Research												

Doing with current resources
Need incremental increase
Requires substantial add.'l resources

Climate Goal by Program

Goal Team Program	End-to-End Environmental Data Management Functions											
	Observation Acquisition & Transmission			Scientific Data Management				Archive & Access				Contin-gency Planning
	5-Year Plan*	Maintain & Monitor	Collect & Rescue	5-Year Plan*	Calibrate & Validate	Appropriate Formats	Complete Metadata	5-Year Plan*	Long-term Preservation	Data Discovery	Access / Disseminate	
Climate Observations and Analysis					<				>		<	
Climate Forcing												
Climate Predictions and Projections												
Climate and Ecosystems	>			>				>				
Regional Decision Support				>				>				

Doing with current resources
Need incremental increase
Requires substantial add.'l resources

Weather & Water by Program

Goal Team Program	End-to-End Environmental Data Management Functions											
	Observation Acquisition & Transmission			Scientific Data Management				Archive & Access				Contingency Planning
	5-Year Plan*	Maintain & Monitor	Collect & Rescue	5-Year Plan*	Calibrate & Validate	Appropriate Formats	Complete Metadata	5-Year Plan*	Long-term Preservation	Data Discovery	Access / Disseminate	
Air Quality	>			>				>				
Coast, Estuaries and Ocean				>	>		>					
Hydro	>			>				>				
Local Forecasts				>		<		>			<	
Space Weather												>
Science and Tech Infusion				>				>				
Tsunami	>		>	>				>	>	>	>	

Doing with current resources

Need incremental increase

Requires substantial add.'l resources

Commerce & Transportation by Program

Goal Team Program	End-to-End Environmental Data Management Functions											
	Observation Acquisition & Transmission			Scientific Data Management				Archive & Access				Contingency Planning
	5-Year Plan*	Maintain & Monitor	Collect & Rescue	5-Year Plan*	Calibrate & Validate	Appropriate Formats	Complete Metadata	5-Year Plan*	Long-term Preservation	Data Discovery	Access / Disseminate	
Marine Transportation System												
Geodesy	<									<	<	<
Surface Weather												
Aviation Weather												
Marine Weather												
NOAA Emergency Response												

Doing with current resources

Need incremental increase

Requires substantial add.'l resources